

FUZZY-NEURAL MULTIPLE BONE SEX CLASSIFICATION SYSTEM UTILIZING GENETIC ALGORITHMS

B.C. Merkl, L.N. Smith, M.R. Mahfouz, A.M. Badawi
Center for Musculoskeletal Research, University of Tennessee, Knoxville, TN, USA
e-mail: mmahfouz@utk.edu

Abstract—A bone classification system is developed to classify the sex of an individual given between one and four bone models representing combinations of femur, tibia, L5 lumbar vertebrae, or intramedullary canal models. Several classification methodologies are employed in classifying individual bones including linear methods using Canonical Variate Analysis (CVA), non-linear methods using neural networks (NN), and non-parametric methods using k -nearest neighbor. All of these methods are implemented to run on a set of features developed using Principal Components Analysis (PCA) computed on the entire set of vertices that comprise a bone model. Finally, an overall classifier merges the classification estimates from the individual bone NN classifiers using a fuzzy inference mechanism and is trained on a holdout set of bones using a genetic algorithm.

Keywords – Bone classification, neural networks, fuzzy inference

I. INTRODUCTION AND PREVIOUS WORK

Forensic anthropology was primarily developed as a means of characterizing certain traits of an individual. It was not until further advancements were made that this type of analysis was used in conjunction with the criminal justice system. Since then, the statistics associated with specific bone measurements have been used to determine characteristics such as age, sex, and ethnicity. In the field of bone classification, it is widely accepted that anthropometric standards can vary significantly from one population to the next. Sex classification is an ongoing area of interest among forensic anthropologists, and has recently extended to other disciplines such as paleoanthropology and archaeology. Ideally, a full skeleton is preferred when attempting to identify an unknown body; however, incomplete or damaged remains are most likely to be recovered. Several publications have demonstrated that sex determination can also be achieved using individual bones in either the upper or lower extremities. Studies that attempt to identify sexual dimorphism in the femora, tibiae, patellae, intramedullary canal (hereafter IM canal), and L5 lumbar vertebrae are most relevant to this particular work. Of these five bones, femoral measurements are most common. DiBennardo and Taylor [1] used both simple and multiple discriminant functions to correctly identify 82% of femora in a White North American population. Mall-Graw et al. [2] evaluated a contemporary German population and found that 91.7% of cases could be accurately classified. Schuler-Ellis et al. [3] were able to accurately predict 80% in regards to femoral length, relative to the pelvis.

Similar to the femora, sex determination based upon tibiae measurements are most often calculated using multivariate discriminant function analysis. Miller-Shaivitz [4] could correctly identify 77% of a sample, while Symes and Jantz [5] achieved a classification accuracy of 82-91%. A more recent

survey from Iscan [6] demonstrated that tibiae measurements could correctly predict sexual dimorphism at a rate of 89%. Holland [7] reported an accuracy of 86-95% when measuring 100 tibial condyles in the Hamann-Todd collection. Shortly after, Kieser et al. [8] studied both Caucasoid and Negro datasets in the R.A. Dart collection and were able to correctly identify 84.62-92%. Due to a recent increased interest in sex classification, many researchers have begun to use patellae measurements. One of the first publications to address this topic was O'Connor [9], who was able to correctly classify 82.5% in females and 78.6% in males. Introna-DiVella et al. [10] reached a rate of 83.3%. In a much more recent study, Kemkes-Grottenthaler [11] reported a classification accuracy of 84%; however, when considering the size of the sample, the sexing accuracy dropped to 74-78%. Bidmos et al. [12] investigated a South African population and found an average accuracy of 83% when using measurements of the maximum height and maximum breadth. Little research has been found with regard to the IM canal and L5 lumbar.

II. MATERIALS AND METHODS

A. Model Creation

To create bone models, CT datasets were manually segmented and surface models consisting of triangulated meshes were generated using Amira™. Several different bone types were used to perform analysis of our NN classification method. Table 1 below highlights the bone types for which we have performed analysis and also includes the total number of male and female instances. The models used in this analysis had between 2000 and 5000 uniformly distributed vertices.

B. Feature Extraction

Feature extraction can be separated into two distinct paradigms. First, geometric-based feature extraction has been performed previously on both the femur [13] and patella [14]. For the patella, these consisted mainly of global features (i.e. applied to the whole model). As for the femur, these methods were only applied distally and proximally. It must be noted that these measurements, while calculated automatically, must be defined manually, and thus this method requires a time-consuming effort to design and code algorithms for feature extraction. Instead, a second method is applied involving PCA performed on matched points across all the models of a particular bone type. This analysis was performed on femora, patellae, tibiae, IM canal, and L5 lumbar vertebrae.

TABLE 1
NUMBER OF MODELS USED FOR TESTING

Bone	Females	Males	Total	PCA Features
Femur	62	103	165	11
Patella	98	142	240	49
IM canal	21	37	58	8
L5	5	9	14	13
Tibia	16	23	39	8

To create the homologous point sets requisite of PCA, we use a series of registration and warping techniques to pick corresponding points on all the other bone models in the training set. The process of picking point correspondences on new models to be added to the atlas is “non-trivial” [15]. The matching algorithm that follows uses several well-known techniques of computer vision as well as our novel contribution for final surface alignment. First, the centroids of the template mesh and new mesh are aligned and the template mesh is pre-scaled to match the bounding box dimensions of the template model with that of the new mesh. Second, a rigid alignment of the template mesh to the new mesh is performed using a standard vertex-to-vertex Iterative Closest Point (ICP) algorithm [16]. Third, we perform a general affine transformation without iteration. Finally, a process called Mutual Correspondence Warping (MCW) is utilized as a non-linear method for final surface alignment [17].

This method is automatic and can be used with a majority of bone topologies. In general, this automatic feature extraction method provides a quick and convenient way to perform analysis on new bone types. These PCA coordinates are recorded for each model and are later used for classification. Using a cumulative sum of variance (dictated by the eigenvalues), we choose to record only the eigenvalues that represent a total of 98% of the cumulative variance (the number of PCA features per bone is shown in Table 1).

C. Classification Methods

Classification methods can be broken down into three large categories: linear, nonlinear, and nonparametric. In this paper, we investigate all three types. This section illustrates the use of linear and non-parametric methods. Of all of the possible linear methods we chose to utilize CVA. While this method is optimal for many classes, for the case of only two (male/female), the result is the same as a Linear Discriminate Function (LDF). We chose CVA on the basis of using this method in the future with a greater number of classes (e.g. age, ethnic origin, etc.). Equations (1-3) demonstrate that this method maximizes between-class covariance and minimizes within-class covariance as a linear system:

Where $\bar{\mu}$ is the overall mean of all the individual class means μ_j for g classes. The matrix B is the between-class covariance, Σ is a pooled estimate of within-class covariance, and N is the number of features. CVA is performed by calculating the eigenvectors of $\Sigma^{-1}B$. By sorting the eigenvectors using the corresponding eigenvalues arranged in descending order, an

$$\bar{\mu} = \frac{1}{g} \sum_{j=1}^g \mu_j \quad (1)$$

$$B = \frac{1}{g-1} \sum_{j=1}^g (\mu_j - \bar{\mu})(\mu_j - \bar{\mu})^T \quad (2)$$

$$\Sigma = \frac{1}{N-1} \sum_{c=1}^g \left\{ \sum_{i=1}^{n_c} (x_{c,i} - \mu_c)(x_{c,i} - \mu_c)^T \right\} \quad (3)$$

ordered set of orthogonal vectors is achieved. The original data is then projected onto these vectors. The first eigenvector provides the maximum class separation, and the eigenspace given by the first two eigenvectors provides the second most class separation, etc. When using CVA with only two classes, only one vector is needed to maximize class separation; thus, projecting the data against additional vectors does not increase separation.

Nonparametric methods include rank transforms and nearest-neighbor techniques, which have applications in density estimation and classification. We use k-Nearest Neighbor (k-NN) classification on the models after they have been projected into CVA space. This projection ensures that the resultant classes are maximally separated so that when the k-NN algorithm is applied the most appropriate class is chosen. By using subspace methods such as PCA and CVA the search time for the nearest neighbors is reduced.

For purely linear classification, once the PCA features have been projected into CVA-space, the data can be classified using a Minimum Distance Classifier (MDC). In this method, we are using merely the PCA coordinates as features, though geometric features could also be used. This procedure is performed as a jackknife test [18], where one model is left out of the dataset; PCA is performed on the remaining models. Next, CVA is calculated on the PCA variables and the largest canonical eigenvector is saved. The PCA coordinates are projected onto the CVA vector resulting in scalar values for each model. These scalar values are averaged class-wise to find the class mean for both males and females. Subsequently the left out model is projected onto the PCA coordinates and then onto the largest canonical vector. The left out model is then classified according to the closest class mean. This method is effectively a MDC. In the case of k-NN classification, the last step is replaced by finding the k-nearest neighbors in CVA-space. Each of these nearest neighbors votes for the class that they represent with the unknown model being classified according to plurality rule.

D. Jackknife NN Optimization

For nonlinear classification we have chosen an implementation of NN that performs well on a subset of the PCA features used in the linear method. After performing PCA, we then created feed-forward back propagation NN. The PCA coordinates for each model are presented to the network on the input layer consisting of n neurons. There is one hidden layer of m neurons (with tansig thresholding functions) and one output neuron that can vary between zero and one (logsig thresholding function) where 0 output is considered female and 1 is considered male. We present

a novel method by which overall network performance can be evaluated without regard to a particular training set.

Before outlining the algorithm of this method, a discussion of the relevant assumptions is in order. The network is trained using 2/3 of the available dataset and the network is tested on the remaining 1/3 of the dataset. Several variables are intended to be optimized and thus will be discussed independent of their nominal values. These variables are the learning rate (η), momentum (α), number of hidden layer neurons (m), and final Mean Squared Error (MSE) that defines convergence. The maximum number of epochs is fixed at 5000 for all bone types, while the repetition number (R) as described below is fixed at 5.

Jackknife Neural Network Algorithm

1. Randomize the order of the input data.
2. Divide the data into sections according to the testing fraction. In our case the data is divided into thirds.
3. Each section is successively used as the testing portion, with the remaining sections used as the training portion.
4. By evaluating those left out, we now have a NN output value for each model in the dataset.
5. Steps 1-4 are repeated a number of times to arrive at an unbiased estimate of the 'average' NN output for every model. In our specific case, steps 1-4 are repeated $R=5$ times, resulting in 5 outputs per model, with a total of $R*3=15$ Neural Networks being created.
6. If a new bone is to be classified, it is input to each of the $R*3$ NNs. A threshold is set to $R*3/2$. If the sum of the 15 networks is above the threshold then the bone is classified as male, otherwise female.

The steps outlined above result in multiple outputs for every model in the dataset. In general, these outputs can be summed and applied to a threshold of $R/2$ to yield an unbiased estimate of NN performance on the dataset. Since there are a number of free parameters, we define an objective function in terms of these free parameters with the purpose of maximizing the unbiased NN classification percentage. By using a genetic algorithm, we optimized the parameters such that the classification percentage of the NN is maximized across the entire dataset.

E. Über Classifier using Fuzzy Inference

In general, for each patient we may have from one to four types of bone models. We seek a method that can combine the jackknife NN data from each available bone into a cohesive classification decision for each patient. Since bone models are often missing for a specific person, the über (overall) classifier must work with missing values as well as the uncertainty in regards to the NN output (Fig 1).

Our system is a Sugeno-type [19] fuzzy inference mechanism that can take multiple inputs and using a set of fuzzy rules determine the sex of the bones. Since each of the NN outputs a total of R values for each model, a simple weighting and thresholding is not appropriate. An observed membership function for a particular bone to a specific sex is approximated by a Gaussian distribution given the mean and standard deviation from the output of the NN sub-classifier.

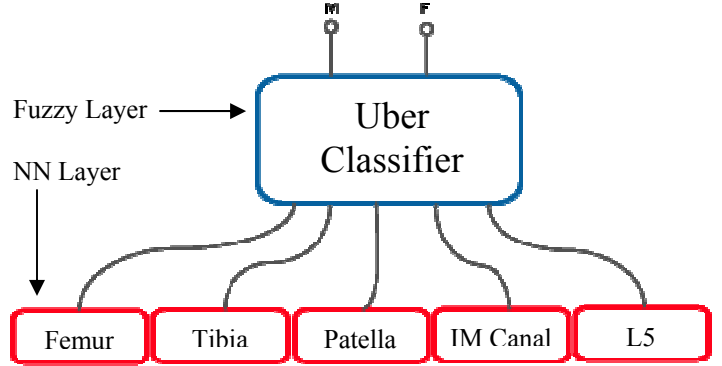


Fig. 1. Über classifier depicting the sub-classifiers which are NN based on the jackknife procedure. The outputs from the bone types is then aggregated using fuzzy inference to produce a consensus value for sex

Using the intersections of the observed membership function with the bone-specific membership function for each sex, a maximum membership value of m_f and m_m is found for the female rule and the male rule respectively. The heights, m_f and m_m , are used to determine the overall height of the fuzzy output function. The multiple fuzzy outputs from each bone type are combined using the union operator; thus, this is a min-max composition. Instances of missing bones do not decrease the performance of the über classifier directly. It only provides the fuzzy inference mechanism with less information.

The last step is to defuzzify the output. This is obtained by finding the center of gravity (CG) of the fuzzy output function. When the CG is above a threshold value of 0.5, the bones are classified as male, otherwise female.

Since it is ambiguous which set of rules is the best, we optimized the rule-set using another genetic algorithm. We framed this training session as a holdout experiment by training the fuzzy rules and fuzzy outputs on one half of the possible data and then testing on the remaining half of the data. The membership functions are defined as a sigmoidal function with center T and squashing K . The output functions will terminate at 1.0 on their respective sides (for each male and female) and will intersect the zero membership line at value b . Thus for each bone type there is a T , K , and b for each sex, bringing the total number of rules to 6 for each bone type. There are 5 bone types so the dimensionality of the optimization algorithm is 30.

III. RESULTS

A. Linear and Non-Parametric Methods

The accuracy of projecting all PCA coordinates onto one CVA direction combined with MDC to determine class is shown in Table 2.

TABLE 2
CVA CLASSIFICATION PERCENTAGES (PURE LINEAR)

Bone Type	Female (%)	Male (%)	Overall (%)
Femur	96.774	100	98.788
Patella	87.755	89.437	88.75
IM canal	80.952	83.784	82.759
L5	100	55.556	71.429
Tibia	100	91.304	94.872

The accuracy for k=3 k-NN algorithm is shown in Table 3. The results are identical to those above, for other k values of 1 and 5 there was a significant drop-off in classification.

TABLE 3
CVA K-NN CLASSIFICATION PERCENTAGES, K=3

Bone Type	Female (%)	Male (%)	Overall (%)
Femur	96.774	100	98.788
Patella	86.735	89.437	88.333
IM canal	85.714	83.784	84.483
L5	100	55.556	71.429
Tibia	100	91.304	94.872

B. Neural Networks as Individual Classifiers

Using the jackknife NN technique as outlined above, each bone type can be quantified with regard to its total ability to classify bones. Table 4 below characterizes our results using this technique on various bone types.

TABLE 4
CLASSIFICATION PERCENTAGES USING THE JACKKNIFE NN
PROCEDURE AGAINST EACH BONE TYPE INDIVIDUALLY

Bone Type	Total	Features	%
Femur	165	11	92.121
Patella	240	49	85.833
IM canal	58	8	74.138
L5	14	13	64.286
Tibia	39	8	94.872

C. Über Classifier

After running the genetic algorithm over 50% of data in the training phase, we achieved results of 88%-92% and 87-91% overall, which was closely mirrored in the testing phase. The genetic algorithm was run with a population of 100 for 200 generations and reached convergence. Table 5 shows the percent classification according to how many bones were present for the person. More bones clearly results in a trend of better classification. When four bones are present, a L5 lumbar is always included, which has the effect of reducing the overall classification percentage.

IV. DISCUSSION

Overall our results for both linear and nonlinear methods have proved to be very similar to or exceed those in published work. We have performed testing on two anatomical structures (IM canals and L5 Lumbar Vertebrae), which have not been seen in the literature. One additional test needs to be performed on the NN – Über Classifier system with all of the Principal Components. It seems that some of the lesser Principal Components, being left out, reduced the separability of the classes.

TABLE 5
PERCENT CLASSIFICATION AS A FUNCTION
OF THE NUMBER OF BONE MODELS PRESENT

Number of Bones	Number of Instances	%
1	203	82.76
2	84	92.86
3	39	94.87
4	7	85.71

REFERENCES

- [1] R. DiBennardo, J.V. Taylor, "Sex assessment of the femur: A test of a new method," *Am. J. Phys. Anthropol.*, vol. 50, pp. 635-638, 1979.
- [2] G. Mall, M. Graw, K. Gehring, M. Hubig, "Determination of sex from femora," *Forensic Sci. Int.*, vol. 113, pp. 315-321, 2000.
- [3] F.P. Schuller-Ellis, L.C. Hayek, D.J. Schmidt, "Determination of sex with discriminant analysis of new pelvic bone measurements: Part II", *J. Forensic Sci.*, vol. 30 (1), pp. 78-185, 1985.
- [4] M.Y. Isçan, P. Miller-Shaivitz, "Determination of sex from the tibia," *Am. J. Phys. Anthropol.*, vol. 64, pp. 53-57, 1984.
- [5] S.A. Symes, R.L. Jantz, "Discriminant function sexing of the tibia," *Am. Acad. of Forensic Sci.*, 1983.
- [6] M.Y. Isçan, M. Yoshino, S. Kato, "Sex determination from the tibia: Standards for contemporary Japan," *J. Forensic Sci.*, vol. 39, pp. 785-792, 1994.
- [7] T.D. Holland, "Sex assessment using the proximal tibia," *Am. J. Phys. Anthropol.*, vol. 85, pp. 221-227, 1991.
- [8] J.A. Kieser, J. Moggi-Cecchi, H.T. Groeneveld, "Sex allocation of skeletal material by analysis of the proximal tibia," *Forensic Sci. Int.*, vol. 56, pp. 29-36, 1992.
- [9] W.G. O'Connor, "The dimorphic sesamoid: differentiating the patella of females and males by height, width and thickness measurements," Master's thesis, Department of Anthropology-University of South Carolina, 1996.
- [10] F. Introna Jr, G. Di Vella, C.P. Campobasso, "Sex determination by discriminant function analysis of patella measurements," *Forensic. Sci. Int.*, vol 95, pp. 39-45, 1998.
- [11] A. Kemkes-Grottenthaler, "Sex determination by discriminant analysis: an evaluation of the reliability of patella measurements," *J. Forensic Sci.*, vol. 147, pp. 129-133, 2005.
- [12] M.A. Bidmos, N. Steinberg, K.L. Kuykendall, "Patella measurements of South African whites as sex assessors," *HOMO*, vol. 56, pp. 69-74, 2005.
- [13] Mahfouz, M.R., Booth, R.E., Jr, Argenson, J.N, Merkl, B.C., Abdel Fatah, E.E., Kuhn, M.J., "Analysis of Variation of Adult Femora Using Sex-Specific Statistical Atlases", 7th Intl Symp CMBBE, Antibes, Cote d'Azur, France, 2006.
- [14] Mahfouz, M.R., Badawi, A.M., Merkl, B.C., Abdel Fatah, E.E., Pritchard, E.R., Kesler, K., Moore, M., Jantz, R., "3D Statistical Shape Models of Patella for Sex Classification", 28th Annual IEEE EMBS, New York, 2006.
- [15] Lorenz C., Krahnstoever N., "Generation of Point-Based 3D Statistical Shape Models for Anatomical Objects," *Computer Vision and Image Understanding*, vol. 77, pp. 175-181, 2000.
- [16] P. J. Besl and N. McKay. "A method for registration of 3-D shapes," *IEEE PAMI*, vol. 14 (2), pp. 239-256, Mar 1992.
- [17] Merkl B.C., Mahfouz M.R., "Unsupervised Three-Dimensional Segmentation of Medical Images Using an Anatomical Bone Atlas," *Int. Conf. Biomed. Engr.*, Singapore, 2005.
- [18] Berg, E., Mahfouz, M.R., Debrunner, C., and Hoff, W., "A 2D Fourier Approach to Deformable Model Segmentation of 3D Medical Images," *Proc. of IEEE Int'l Symposium on Biomedical Imaging*, Arlington VA, April 2004.
- [19] Sugeno, M., Asai, K., and Terano, T., "Applied Fuzzy Systems. AP Professional", Cambridge, 1989.